Density Estimation in Several Populations With Uncertain Population Membership

Yanyuan MA, Jeffrey D. HART, and Raymond J. CARROLL

We devise methods to estimate probability density functions of several populations using observations with uncertain population membership, meaning from which population an observation comes is unknown. The probability of an observation being sampled from any given population can be calculated. We develop general estimation procedures and bandwidth selection methods for our setting. We establish large-sample properties and study finite-sample performance using simulation studies. We illustrate our methods with data from a nutrition study.

KEY WORDS: Cross-validation; Least squares; Plug-in method; Uncertain sample.

1. INTRODUCTION

Suppose that we observe S, which is known to come from one of P subpopulations with densities $f_1, \ldots, f_P$. We are interested in the nonparametric estimation of each of these P densities. Our problem differs from the usual density estimation framework, because each observation on S is made without knowing from which subpopulation it came. Instead, we observe $\mathbf{Q} = (q_1, \ldots, q_P)$ with $\sum_{j=1}^{P} q_j = 1$, where q_j is an estimate of the probability that S comes from subpopulation j, $j = 1, \ldots, P$. Thus subpopulation membership is missing, but membership probabilities are known or can be estimated.

Simple examples of this problem arise quite naturally in nutritional epidemiology. For example, suppose that we are interested in a quantitative trait S, such as body mass index or triglyceride level, among subpopulations formed by the $P = 3$ tertiles of usual dietary behavior, for example, the tertiles of caloric intake or the levels of the Healthy Eating Index-2005 (HEI-2005, *http://www.cnpp.usda.gov/HealthyEatingIndex.htm*). Unfortunately, usual dietary intake cannot be measured exactly, because the instruments used either focus only on short-term diet, such as 24-hour recalls, or use a food frequency questionnaire. Generally, in nutritional epidemiology and surveillance, information about usual intake is available from repeated administration of the instrument followed by a measurement error analysis that estimates the distribution of usual intake, from which the probability of membership in each tertile can be estimated. Some examples of complex measurement error analysis have been given by Nusser, Fuller, and Guenther (1997), Dodd et al. (2006), Johnson et al. (2007), and Sinha et al. (2010), among others. Of course, our problem can arise in many other contexts as well. These general methods result in $n^{1/2}$-consistent estimation of the membership probabilities, which means that a nonparametric density estimator based on estimated probabilities will have the same asymptotic distribution as in the case where membership probabilities are known.

Yanyuan Ma is Professor (E-mail: *ma@stat.tamu.edu*), Jeffrey D. Hart is Professor (E-mail: *hart@stat.tamu.edu*), and Raymond J. Carroll is Distinguished Professor (E-mail: *carroll@stat.tamu.edu*), Department of Statistics, Texas A&M University, College Station, TX 77843-3143. Ma's research was supported by the National Science Foundation grant DMS-090634 and National Institute of Health grant R01-NS 073671. Carroll's research was supported by National Cancer Institute grant R37-CA057030 and National Science Foundation Instrumentation grant 0922866. This publication is based in part on work supported by King Abdullah University of Science and Technology (award KUS-CI-016-04).

Let m_i denote the unknown subpopulation from which S_i comes, $i = 1, \ldots, n$. Our data are then $(S_i, \mathbf{Q}_i)$, $i = 1, \ldots, n$, where $q_{ij} = \mathrm{pr}(m_i = j), j = 1, \ldots, P$. At a fixed $\mathbf{Q}_i$, the marginal density function of S_i is the mixture $f_S(s) = \sum_{j=1}^{P} q_{ij} f_j(s)$. Our goal is to estimate $\mathbf{f}(s) = \{f_1(s), \ldots, f_P(s)\}^{\mathrm{T}}$. Here, as is motivated in our example, $\mathbf{Q}_i$ is a multivariate continuous random variable. This is in contrast to the setting of Ma and Wang (2010), where $\mathbf{Q}_i$ is discrete, and thus density estimation becomes straightforward. Specifically, in their problem, $\mathbf{Q}_i$ can take on only finitely many, say m, vector values $\mathbf{u}_1, \ldots, \mathbf{u}_m$. Thus the n observations can be grouped into m groups according to their $\mathbf{Q}_i$ values. Then a classical nonparametric density estimation within each group is performed to obtain m density estimates $\widehat{\mathbf{g}}_1(s) = \mathbf{u}_1 \widehat{\mathbf{f}}(s), \ldots, \widehat{\mathbf{g}}_m(s) = \mathbf{u}_m \widehat{\mathbf{f}}(s)$, and $\widehat{\mathbf{f}}(s)$ is recovered through $\widehat{\mathbf{f}}(s) = (\sum_{i=1}^{m} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}})^{-1} \{\sum_{i=1}^{m} \mathbf{u}_i \widehat{\mathbf{g}}_i(s)\}$. This simple treatment does not apply when $\mathbf{Q}_i$ is continuous, because then we could have $m = n$, and each group would only have one observation.

We also point out connections and differences between our problem and more widely studied clustering problems. In both cases, each observation is assumed to be drawn randomly from one of several populations. However, in clustering the probability density functions of the populations are often clearly differentiated (Mallapragada, Jin, and Jain 2010), whereas in our problem these functions do not even have to be different. In addition, in clustering, the likelihood that an observation belongs to a specific population is unknown, and identifying the population to which a particular observation belongs is of interest. However, in our problem, the probability that an observation belongs to a specific population is either known or can be estimated at a root-n rate. For these reasons, our problem is substantially different from both the clustering problem and classical mixture models.

The rest of the article is organized as follows. In Section 2 we derive a general method for constructing a class of nonparametric estimators. We also derive the most efficient estimator in this class. In Section 3 we establish asymptotic properties of the estimators. We report simulation experiments in Section 4, and use the proposed estimators to analyze a real data example in Section 5. We conclude with a discussion in Section 6, and provide technical details are provided in the Appendix.

Journal of the American Statistical Association
September 2011, Vol. 106, No. 495, Theory and Methods
DOI: 10.1198/jasa.2011.tm10798

2. ESTIMATION

2.1 Basic Estimator

We propose a family of kernel-based weighted least squares estimators,

$$\widehat{f}_j(s) = \mathbf{e}_j^{\mathrm{T}}\left(n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right)^{-1} n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i K_{h_j}(S_i - s) \quad (1)$$

for $j = 1, \ldots, P$. Here w_i is an arbitrary weight, h_j is a bandwidth, and $\mathbf{e}_j$ is a vector of length P with 1 as the jth entry and 0s elsewhere. For any bandwidth h, $K_h(s) \equiv K(s/h)/h$, where K is a kernel function. Among different choices of weights, the optimal ones are $w_i = \{\mathbf{q}_i^{\mathrm{T}}\mathbf{f}(s)\}^{-1}$, $i = 1, \ldots, n$.

Estimator (1) is motivated by the following considerations. Given $\mathbf{q}_i$, the probability of $S_i < s$ for any fixed s can be written as $E\{I(S_i < s)|\mathbf{Q}_i = \mathbf{q}_i\} = \mathbf{q}_i^{\mathrm{T}}\mathbf{F}(s)$, where $\mathbf{F}(s)^{\mathrm{T}} = (F_1(s), \ldots, F_P(s))$ and F_j is the cumulative distribution function (cdf) corresponding to subpopulation j. Viewing this as a linear regression problem yields $I(S_i < s) = \mathbf{q}_i^{\mathrm{T}}\mathbf{F}(s) + e_i$, where e_i satisfies $E(e_i|\mathbf{q}_i) = 0$ and $\mathrm{var}(e_i|\mathbf{q}_i) = \mathbf{q}_i^{\mathrm{T}}\mathbf{F}(s)\{1 - \mathbf{q}_i^{\mathrm{T}}\mathbf{F}(s)\}$ (Ma and Wang 2010). Thus a weighted least squares estimator for the vector of cdfs is given by

$$\widetilde{\mathbf{F}}(s) = \left(n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right)^{-1} n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i I(S_i < s). \quad (2)$$

Differentiating these step functions yields

$$\widehat{\mathbf{f}}(s) = \left(n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right)^{-1} n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i \delta(S_i - s), \quad (3)$$

where $\delta(\cdot)$ denotes the Dirac delta function. Equation (3) can be viewed as an empirical density estimator, inasmuch as it places a point mass at each S_i. To obtain a smooth density estimate, we replace the point mass at S_i by a kernel centered at S_i. Considering that different components of $\mathbf{f}(s)$ can have different smoothness, we further allow component-specific bandwidths, leading us to (1).

For readers familiar with the classic technique of linking a distribution function estimator $\widehat{F}$ and a density estimator $\widehat{f}$ via the convolution $\widehat{f}(s) = \int K_h(s-t)\, d\widehat{F}(t)$, the estimator in (1) also can be obtained as follows:

$$\begin{aligned}
\widehat{f}_j(s) &= \int K_{h_j}(s-t)\, d\{\mathbf{e}_j^{\mathrm{T}}\widetilde{\mathbf{F}}(t)\} \\
&= \int K_{h_j}(s-t)\, d\left\{\mathbf{e}_j^{\mathrm{T}}\left(n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right)^{-1} \times n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i I(S_i < t)\right\} \\
&= \mathbf{e}_j^{\mathrm{T}}\left(n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right)^{-1} \times n^{-1}\sum_{i=1}^{n} w_i q_i \int K_{h_j}(s-t)\, d\{I(S_i < t)\} \\
&= \mathbf{e}_j^{\mathrm{T}}\left(n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right)^{-1} n^{-1}\sum_{i=1}^{n} w_i\mathbf{q}_i K_{h_j}(s - S_i).
\end{aligned}$$

2.2 Bandwidth Estimation

We now study the issue of bandwidth selection for h_j. Because the estimation procedure is a weighted kernel estimator, leave-one-out least squares cross-validation (CV) (Bowman 1984) is a candidate method. Let $\widehat{f}_j(\cdot; h)$ denote estimator (1) based on bandwidth h. The integrated squared error (ISE) of $\widehat{f}_j$ is $\mathrm{ISE}(\widehat{f}_j) = \int\{\widehat{f}_j(s; h_j) - f_j(s)\}^2\, ds$. Minimizing $\mathrm{ISE}(\widehat{f}_j)$ with respect to h is equivalent to minimizing

$$\mathrm{ISE}(\widehat{f}_j) - \int f_j^2(s)\, ds = \int \widehat{f}_j^2(s; h_j)\, ds - 2\int \widehat{f}_j(s; h_j) f_j(s)\, ds. \quad (4)$$

CV provides an estimator of (4). For $i = 1, \ldots, n$, let $\widehat{f}_{j,-i}(\cdot; h_j)$ be the weighted least squares density estimator using all of the observations except S_i. We define the CV bandwidth to be the minimizer $\widehat{h}_j$ of the following criterion:

$$\begin{aligned}
\mathrm{CV}_j(h) &= \int \widehat{f}_j^2(s; h)\, ds - 2\int \widehat{f}_{j,-i}(s; h)\, d\mathbf{e}_j^{\mathrm{T}}\widetilde{\mathbf{F}}(s) \\
&= \int \widehat{f}_j^2(s; h)\, ds \\
&\quad - 2\sum_{i=1}^{n}\widehat{f}_{j,-i}(S_i; h)\mathbf{e}_j^{\mathrm{T}}\left(n^{-1}\sum_{j=1}^{n} w_j\mathbf{q}_j\mathbf{q}_j^{\mathrm{T}}\right)^{-1} n^{-1} w_i\mathbf{q}_i.
\end{aligned}$$

In the usual density estimation context, CV is known to have relatively large variance. To overcome this problem, we consider the indirect CV (ICV) idea proposed by Savchuk, Hart, and Sheather (2010). This method uses different kernels for bandwidth selection and density estimation, which allows the selection of a kernel that, although perhaps not suitable for density estimation, is very good for CV.

Specifically, we use the kernel function $K_{\mathrm{cv}}(x) = (1 + \alpha)\phi(x) - \alpha/\sigma\phi(x/\sigma)$ in the foregoing CV procedure, where $\alpha = 2.42$, $\sigma = \max(5.06, 0.149n^{3/8})$, and ϕ is the standard normal density function. But because K_{cv} is known to perform poorly for density estimation, we use a quartic kernel as our K in the actual estimation of the density function. The final bandwidth for the quartic kernel density estimate of the jth subpopulation density is obtained by first calculating the ICV bandwidth $\widehat{b}_j$, and then rescaling it via

$$\widetilde{h}_j = \widehat{b}_j\left\{\frac{\int K^2(x)\, dx\{\int x^2 K_{\mathrm{cv}}(x)\, dx\}^2}{\int K_{\mathrm{cv}}^2(x)\, dx\{\int x^2 K(x)\, dx\}^2}\right\}^{1/5}.$$

Savchuk, Hart, and Sheather (2010) established that in the standard density estimation setting, $\widetilde{h}_j$ has relative error converging to 0 at the rate $n^{-1/4}$, in contrast to the $n^{-1/10}$ rate of the classical CV. They also pointed out that CV is known to work well in some difficult estimation problems, such as when the density is not smooth or has multiple sharp peaks. Thus an intuitive motivation for ICV is that by using a kernel (K_{cv}) that poorly estimates the density, the estimation problem becomes difficult and so CV works better.

Another possibility is to use a plug-in type of bandwidth selector. Toward this end, we use the asymptotic bias and variance results derived in the next section, and minimize the mean

ISE (MISE). To simplify notation, we write $\sigma_K^2 = \int u^2 K(u)\,du$, $C_K = \int K^2(u)\,du$ and

$$c_j = \mathbf{e}_j^{\mathrm{T}}\left(n^{-1}\sum_{i=1}^n w_i \mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right)^{-1}\left\{n^{-1}\sum_{i=1}^n w_i^2 \mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right\} \times \left(n^{-1}\sum_{i=1}^n w_i \mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right)^{-1}\mathbf{e}_j, \qquad j = 1,\ldots,P.$$

Thus we minimize $\int\{h^2\sigma_K^2 f_j''(s)\}^2\,ds/4 + C_K c_j/(nh)$, which leads to the first-order optimal bandwidth

$$h_{j,\mathrm{opt}} = \left(\frac{C_K c_j}{\sigma_K^4 n}\right)^{1/5}\left\{\frac{1}{\int f_j''(s)^2\,ds}\right\}^{1/5}.$$

The only unknown quantity in the foregoing expression is $\int f_j''(s)^2\,ds$, which can be estimated using

$$\widehat{f}_j''(s) = \mathbf{e}_j^{\mathrm{T}}\left(n^{-1}\sum_{i=1}^n w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right)^{-1} \times n^{-1}\sum_{i=1}^n w_i \mathbf{q}_i \widetilde{h}^{-3}K''\{(S_i - s)/\widetilde{h}\}, \qquad (5)$$

where $\widetilde{h}$ is a new bandwidth. There is an obvious similarity between the optimal bandwidth form $h_{j,\mathrm{opt}}$ and the usual plug-in optimal bandwidth. In addition, estimation of the second derivative $\widehat{f}_j''(s)$ in (5) differs from the usual form only by the weight $\nu_{ji} \equiv \mathbf{e}_j^{\mathrm{T}}(n^{-1}\sum_{i=1}^n w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}})^{-1}n^{-1}w_i\mathbf{q}_i$. This allows us to adapt the idea of the Sheather–Jones plug-in method (Sheather and Jones 1991) to devise a suitable subsample plug-in method. Specifically, the adaptation entails modifying three quantities in their procedure. The first modification is to replace the sample size with an approximate subsample size $n\overline{\mathbf{q}}_j$. The second modification is to define an appropriate estimate of the interquartile range for the jth subsample. Let $s_{(1)} < \cdots < s_{(n)}$ be the ordered observations, and for $i = 1,\ldots,n$ let $\omega_{(i)j}$, $j = 1,\ldots,P$, denote the weights associated with $s_{(i)}$. Our estimate of the interquartile range for subpopulation j is $s_{(k)} - s_{(\ell)}$, where (k,ℓ) satisfies $\sum_{i\le k}\omega_{(i)j} \ge 3/4\sum_{i=1}^n \omega_{(i)j} > \sum_{i\le k-1}\omega_{(i)j}$ and $\sum_{i\le\ell}\omega_{(i)j} \le 1/4\sum_{i=1}^n\omega_{(i)j} < \sum_{i\le\ell+1}\omega_{(i)j}$. The third modification is to replace the double summation in the approximation of two squared integration terms (Sheather and Jones' $\widehat{S}_D$ and $\widehat{T}_D$, respectively) with a weighted version using the ν_{ji} weights. These modifications ensure that the equation solved when each observation belongs to a subpopulation with a given probability is asymptotically equivalent to the equation solved in a case with a known population membership.

2.3 A Locally Efficient Class

Ma and Wang (2010) considered the case where $\mathbf{Q}_i$ is discrete and derived a class of semiparametric estimators of the distribution function. In brief, they took a proposed density $\mathbf{f}^*(s)$ for $\mathbf{f}(s)$ and derived estimators for $\mathbf{F}(s)$ that are efficient if the proposed model is correct and consistent otherwise.

Define

$$\mathbf{A}^*(s) = n^{-1}\sum_{i=1}^n\left\{\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}/(\mathbf{q}_i^{\mathrm{T}}\mathbf{f}^*(s))\right\},$$

a sample average version of $\mathbf{A}_0^*(s) \equiv \int\cdots\int\{\mathbf{q}\mathbf{q}^{\mathrm{T}}g_Q(\mathbf{q})/(\mathbf{q}^{\mathrm{T}}\mathbf{f}^*(s))\}\,d\mathbf{q}$, where g_Q denotes the multivariate density of the P-variate random vector $\mathbf{Q}$, and

$$\mathbf{K}^*(s) = \int I(t<s)\mathbf{A}^*(t)^{-1}\,dt\left\{\int \mathbf{A}^*(t)^{-1}\,dt\right\}^{-1}.$$

Then the estimators of Ma and Wang (2010) are of the form

$$\widehat{\mathbf{F}}(s) = n^{-1}\sum_{i=1}^n \frac{\{I(S_i<s) - \mathbf{K}^*(s)\}\mathbf{A}^*(S_i)^{-1}\mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}}\mathbf{f}^*(S_i)} + \mathbf{K}^*(s)\mathbf{1}_p, \qquad (6)$$

where $\mathbf{1}_p$ is a length p vector with each component 1.

As described in Section 2.1, any distribution function estimator can be modified to a corresponding density estimator. Thus, starting with estimators in the class (6), we can define the class of density estimators

$$\begin{aligned}
\widehat{f}_j(s) &= \mathbf{e}_j^{\mathrm{T}} n^{-1}\sum_{i=1}^n\left[K_{h_j}(S_i - s) - \mathbf{A}^*(s)^{-1}\left\{\int \mathbf{A}^*(t)^{-1}\,dt\right\}^{-1}\right] \\
&\quad \times \mathbf{A}^*(S_i)^{-1}\mathbf{q}_i\big/(\mathbf{q}_i^{\mathrm{T}}\mathbf{f}^*(S_i)) \\
&\quad + \mathbf{e}_j^{\mathrm{T}}\mathbf{A}^*(s)^{-1}\left\{\int \mathbf{A}^*(t)^{-1}\,dt\right\}^{-1}\mathbf{1}_p \\
&= \mathbf{e}_j^{\mathrm{T}} n^{-1}\sum_{i=1}^n \frac{\mathbf{A}^*(S_i)^{-1}\mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}}\mathbf{f}^*(S_i)}K_{h_j}(S_i - s) \\
&\quad + \mathbf{e}_j^{\mathrm{T}}\mathbf{A}^*(s)^{-1}\left\{\int \mathbf{A}^*(t)^{-1}\,dt\right\}^{-1} \\
&\quad \times n^{-1}\sum_{i=1}^n\left\{\mathbf{1}_p - \frac{\mathbf{A}_0^*(S_i)^{-1}\mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}}\mathbf{f}^*(S_i)}\right\} \\
&\quad + \mathbf{e}_j^{\mathrm{T}}\mathbf{A}^*(s)^{-1}\left\{\int \mathbf{A}^*(t)^{-1}\,dt\right\}^{-1} \\
&\quad \times n^{-1}\sum_{i=1}^n\left\{\frac{\mathbf{A}_0^*(S_i)^{-1}\mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}}\mathbf{f}^*(S_i)} - \frac{\mathbf{A}^*(S_i)^{-1}\mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}}\mathbf{f}^*(S_i)}\right\}. \qquad (7)
\end{aligned}$$

Inspecting the three summands in (7) is of interest. The first summand is a weighted average of the kernels $K_{h_j}(S_i - s)$, and thus is $O_p(1)$. Because

$$\begin{aligned}
E\left\{\mathbf{1}_p - \frac{\mathbf{A}_0^*(S_i)^{-1}\mathbf{Q}_i}{\mathbf{Q}_i^{\mathrm{T}}\mathbf{f}^*(S_i)}\right\} &= \mathbf{1}_p - \int \frac{\mathbf{A}_0^*(s)^{-1}\mathbf{q}}{\mathbf{q}^{\mathrm{T}}\mathbf{f}^*(s)}g_Q(\mathbf{q})\mathbf{q}^{\mathrm{T}}\mathbf{f}(s)\,d\mathbf{q}\,ds \\
&= \mathbf{1}_p - \int \mathbf{f}(s)\,ds = 0,
\end{aligned}$$

the second summand is a term of order $O_p(n^{-1/2})$ under condition C3. The third summand is of the same order as $\mathbf{A}_0^*(s) - \mathbf{A}^*(s)$, which is $O_p(n^{-1/2})$ under condition C3. Thus, to first order, the density estimators derived from the family (6) can be written as

$$\widehat{f}_j(s) = \mathbf{e}_j^{\mathrm{T}} n^{-1}\sum_{i=1}^n \frac{\mathbf{A}^*(S_i)^{-1}\mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}}\mathbf{f}^*(S_i)}K_{h_j}(S_i - s). \qquad (8)$$

In Section 3 we show that the estimators in (8) are special cases of the weighted least squares estimators in (1), and thus in applications we need only implement the weighted least squares estimators. This is very different from the situation when estimating $\mathbf{F}(t)$, where the weighted least squares estimators are not locally semiparametric efficient (Ma and Wang 2010).

3. ASYMPTOTIC PROPERTIES

If we let $w_i = \{\mathbf{q}_i^{\mathrm{T}}\mathbf{f}^*(s)\}^{-1}$ in (1), then we see that (1) and (8) are identical except for the arguments inside $\mathbf{f}^*$ and $\mathbf{A}^*$. In (1), both $\mathbf{f}^*$ and $\mathbf{A}^*$ are evaluated at the point of estimation s, whereas in (8), they are evaluated at the observations $S_1, \ldots, S_n$.

The nonparametric density estimators given in (1) are based on kernel estimation, and thus have the usual bias and variance properties. We express the asymptotic results of the weighted least squares estimator in Theorem 1 and state the equivalence relation between the locally efficient estimators and a weighted least squares in Theorem 2. Here and in the sequel, unless stated otherwise, all of the bias and variance results are conditional on $\mathbf{q}_1, \ldots, \mathbf{q}_n$, which is analogous to the familiar practice of conditioning on covariates in regression. The proofs are provided in the Appendix.

Theorem 1. For any arbitrary nonnegative weights w_i, $i = 1, \ldots, n$, under regularity conditions C1–C3 in the Appendix, the weighted least squares estimator $\widehat{f}_j(s)$ in (1) has bias $h_j^2 f_j''(s)\sigma_K^2/2 + O(h_j^4)$. Defining $\mathcal{M}_q = (n^{-1}\sum_{i=1}^n w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}})^{-1}$, for $j = 1, \ldots, P$, the variance is

$$(nh_j)^{-1}\mathbf{e}_j^{\mathrm{T}}\mathcal{M}_q\left\{n^{-1}\sum_{i=1}^n w_i^2\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\mathbf{q}_i^{\mathrm{T}}\mathbf{f}(s)\right\}\mathcal{M}_q\mathbf{e}_j C_K + O(n^{-1}).$$

Remark 1. Using standard matrix manipulation, it is easy to see that among the different choices of w_i's, the one that yields the smallest estimation variance conditional on $\mathbf{q}_1, \ldots, \mathbf{q}_n$ is $w_i = \{\mathbf{q}_i^{\mathrm{T}}\mathbf{f}(s)\}^{-1}$, and the resulting estimation variance is

$$\frac{1}{nh_j}\mathbf{e}_j^{\mathrm{T}}\left[n^{-1}\sum_{i=1}^n\{\mathbf{q}_i^{\mathrm{T}}\mathbf{f}(s)\}^{-1}\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right]^{-1}\mathbf{e}_j C_K + O(1/n).$$

Remark 2. Using the estimator of the distribution function $\widetilde{\mathbf{F}}(s)$ in (2), and again defining $\mathcal{M}_q = (n^{-1}\sum_{i=1}^n w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}})^{-1}$, it is easy to see that the variance is

$$\begin{aligned}\operatorname{var}\{\widetilde{\mathbf{F}}(s)\} &= \operatorname{var}\left\{\left(n^{-1}\sum_{i=1}^n w_i\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\right)^{-1} n^{-1}\sum_{i=1}^n w_i\mathbf{q}_i I(S_i < s)\right\}\\ &= n^{-1}\mathcal{M}_q n^{-1}\sum_{i=1}^n w_i^2\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}\mathbf{q}_i^{\mathrm{T}}\mathbf{F}(s)\{1 - \mathbf{q}_i^{\mathrm{T}}\mathbf{F}(s)\}\mathcal{M}_q.\end{aligned}$$

This shows that the variance of $\widehat{\mathbf{f}}(s)$ is closely linked with that of $\widetilde{\mathbf{F}}(s)$.

Remark 3. In estimating the distribution function $\mathbf{F}(s)$, the optimal weights are $w_i^{-1} = \mathbf{q}_i^{\mathrm{T}}\mathbf{F}(s)\{1 - \mathbf{q}_i^{\mathrm{T}}\mathbf{F}(s)\}$, which differ from the optimal weights in the density estimator $\widehat{\mathbf{f}}(s)$. Thus, although the two estimators are closely linked, a more efficient $\widehat{\mathbf{F}}(s)$ does not necessarily result in a more efficient $\widehat{\mathbf{f}}(s)$.

Theorem 2. Under regularity conditions C1–C3, the locally efficient estimator $\widehat{f}_j(s)$ in (8) is asymptotically equivalent to the weighted least squares estimator in (1) when $w_i^{-1} = \mathbf{q}_i^{\mathrm{T}}\mathbf{f}^*(s)$. Here asymptotic equivalence means that the difference between the two estimators is of smaller order in mean square than the MSE of either estimator.

4. SIMULATIONS

We performed extensive simulation studies to investigate the performance of the proposed procedure. We report here the results of three studies. The first two are designed to illustrate the impact of different weights and different bandwidth selection procedures. The third simulation is designed to mimic the nutrition data analyzed in Section 5.

In the first simulation, we generate data from $P = 2$ populations. Both populations have a standard normal distribution. Each dataset contains 400 observations, where for each observation, the probability that it belongs to the first and second population, $\mathbf{q}_i = (q_{i1}, q_{i2})^{\mathrm{T}}$, is obtained as $q_{i1} = u_1/(u_1 + u_2)$, $q_{i2} = u_2/(u_1 + u_2)$, where u_1, u_2 are generated independently from the uniform $[0, 1]$ distribution. Thus, there are approximately 200 observations from each population, and the two population distributions are identical. We used CV, ICV, and the adapted plug-in methods described in Section 2.2 to select bandwidths, and experimented with both constant weights $w_i = 1$ and the optimal weights $w_i^{-1} = \mathbf{q}_i^{\mathrm{T}}\mathbf{f}(S_i)$. We generated 1000 datasets, and report the resulting values of average ISE for various situations in Table 1. Figure 1 provides plots of pointwise 5%, median, and 95% quantile curves.

The second simulation differs from the first simulation in that the two populations have very different distributions. The first true density is a standard normal distribution with mean 10 and variance 25, whereas the second true density is a Student t distribution centered at 20 with scale parameter 10 and 4 degrees of freedom. We performed the same estimation procedures as in simulation 1, and summarize the results in Table 1 and Figure 2.

Table 1. Average values of ISE for optimal weighted least squares (OWLS) and its constant weight version (OLS), in combination with least squares CV, ICV, and the adapted plug-in method. Each value in the table was computed from 1000 replications and is 100× the actual value

	Simulation 1, $n = 400$		Simulation 2, $n = 400$		Simulation 3, $n = 500$		
	f_1	f_2	f_1	f_2	f_1	f_2	f_3
OLS, ICV	0.73	0.73	0.19	0.07	0.77	0.73	0.60
OLS, CV	1.54	1.65	0.33	0.20	1.76	1.59	1.22
OLS, plug-in	0.82	0.83	0.21	0.08	0.86	0.75	0.60
OWLS, ICV	0.73	0.73	0.19	0.07	0.77	0.73	0.60
OWLS, CV	1.54	1.65	0.31	0.18	1.74	1.58	1.20
OWLS, plug-in	0.82	0.83	0.20	0.08	0.85	0.75	0.59

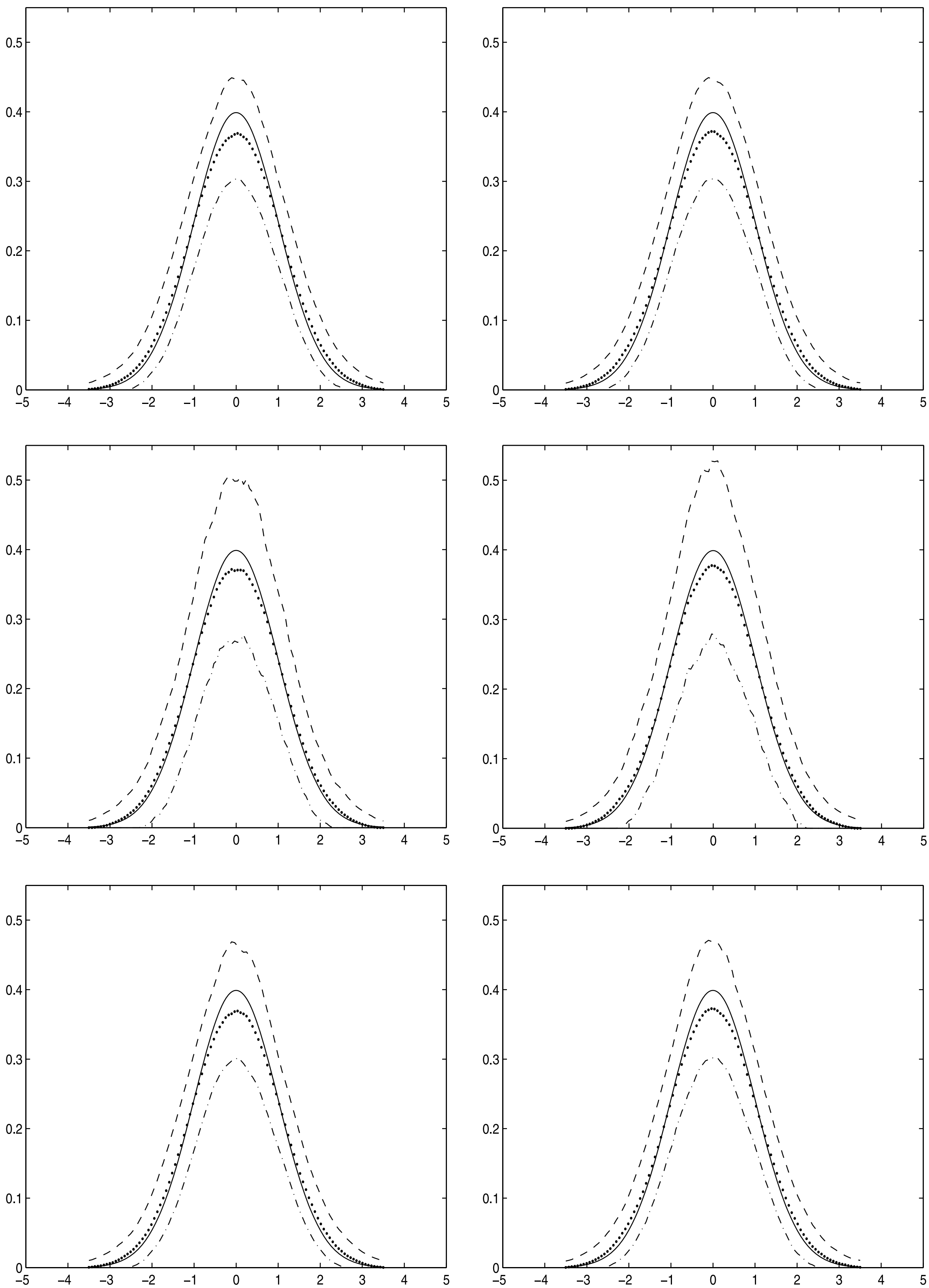

Figure 1. Pointwise quantile curves from simulation 1. In each plot, the solid line is the true density and the other three curves are the median (dotted), 5% (dashed), and 95% (dotted–dashed) quantile curves of all 1000 density estimates in simulation 1. The left and right panels correspond to populations 1 and 2, respectively. From top to bottom, the rows correspond to bandwidth selection by ICV, CV, and the adapted plug-in. All results are based on the constant weight estimators.

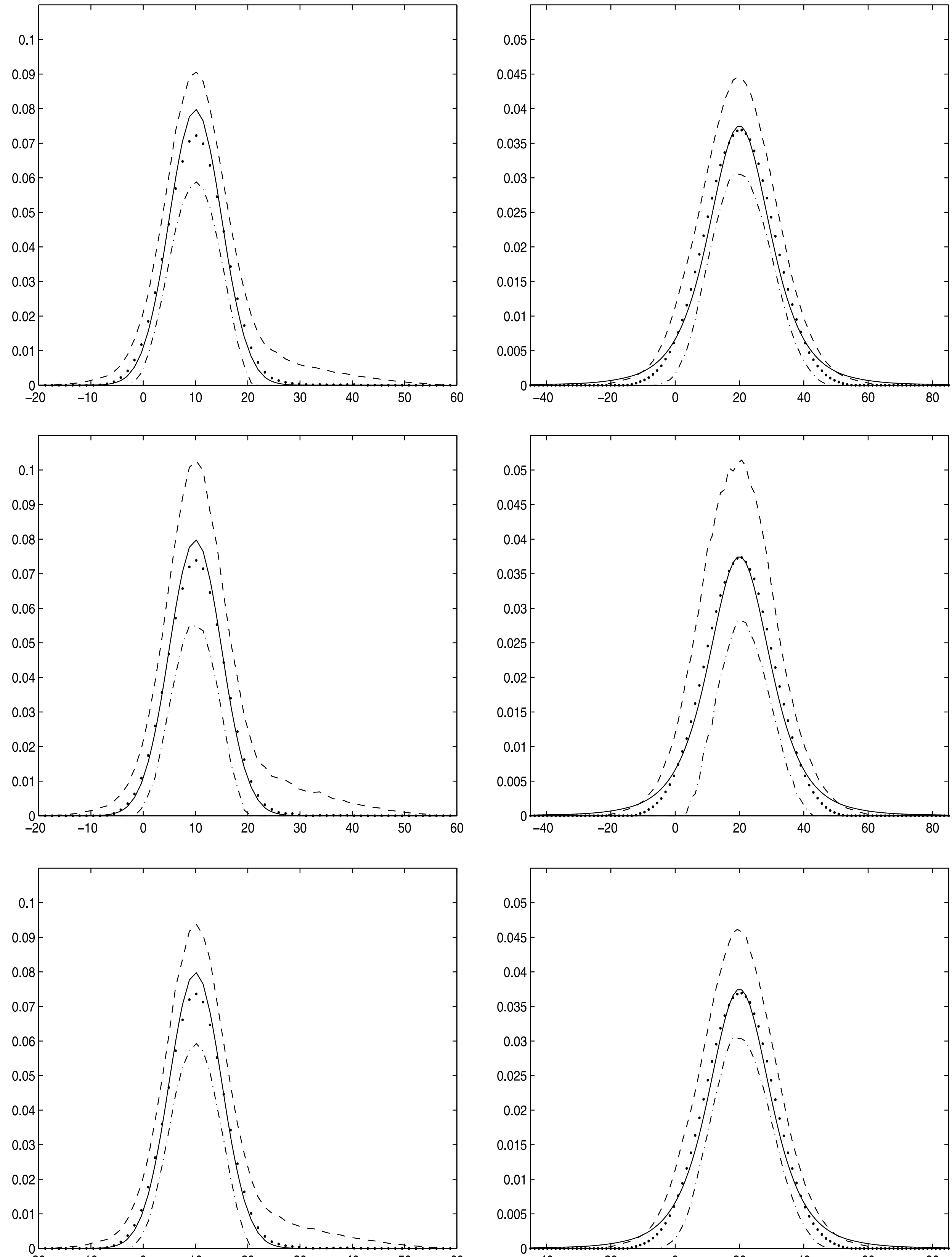

Figure 2. Pointwise quantile curves from simulation 2. In each plot, the solid line is the true density and the other three curves are the median (dotted), 5% (dashed), and 95% (dotted–dashed) quantile curves of all 1000 density estimates in simulation 2. The left and right panels correspond to populations 1 and 2, respectively. The rows, from top to bottom, correspond to bandwidth selection by ICV, CV, and the adapted plug-in. All results are based on the constant weight estimators.

Our third simulation is generated from a mixture of three populations. It has sample size $n = 500$, and each subpopulation is a mixture of two normals, as follows:

$$f_1(x) = (0.98)(1.1)^{-1}\phi\left(\frac{x}{1.1}\right) + (0.02)(0.7)^{-1}\phi\left(\frac{x-4.5}{0.7}\right),$$

$$f_2(x) = (0.80)(1.1)^{-1}\phi\left(\frac{x}{1.1}\right) + (0.20)2^{-1}\phi\left(\frac{x-3}{2}\right),$$

and

$$f_3(x) = (0.60)(1.1)^{-1}\phi\left(\frac{x}{1.1}\right) + (0.40)2^{-1}\phi\left(\frac{x-3}{2}\right).$$

These distributions resemble those in the nutrition example. The probability vector $\mathbf{q}_i$ is also generated to reflect a similar pattern as in the data example, where for each of the three subsamples, $j = 1, 2, 3$, approximately 30% of the samples have a q_{ij} value > 0.9, 60% have a value < 0.1, and the remaining 10% have values evenly distributed throughout $[0, 1]$. The estimation results are presented in Table 1 and Figure 3.

One common observation regarding the results in all of the simulations is that the weight choice w_i does not seem to have a large impact. Because of this, in all the figures, we only show the results of the constant weight estimators. This phenomenon also was observed in distribution function estimation by Ma and Wang (2010). This is certainly an encouraging result from a practical standpoint, in that it justifies using a simple equal weight method.

Our second observation is that the estimation results are quite sensitive to the bandwidth selection method. This finding does not come as a surprise, given that it is usually the case for nonparametric curve estimation. Here we implemented the three selection methods described in Section 2.2, namely CV, ICV, and our adapted version of the Sheather–Jones plug-in method. Properties of the three methods turned out to be similar to those in the familiar single population case. To wit, the CV bandwidth has relatively small bias, but very large variability. The adapted plug-in method reduces variability but is sometimes biased. The ICV method also has relatively small variability, but tends to produce oversmoothed density estimates, as is most evident in f_3 in Figure 3. On the basis of average ISE, however, ICV is the winner among the three methods. ICV produces smaller average ISE compared with CV and the adapted plug-in for all seven densities in the three simulations. Based on these observations, it seems safe to say that CV is not a viable method compared with ICV, whereas the adapted plug-in methods is slightly inferior to ICV.

Finally, to investigate the accuracy of the asymptotic results given in Section 3, we calculated both the sample MISE and the average first-order asymptotic MISE. In each simulation, we integrated (with respect to s) the MSE approximation in Theorem 1 to obtain a first-order approximation of MISE, conditional on $\mathbf{q}_1, \ldots, \mathbf{q}_n$. We then averaged over 1000 simulations to obtain the average first-order asymptotic MISE. As a function of bandwidth, this averaged MISE is a reasonably good approximation to the expected ISE of $\widehat{f}_j$. The results from the three simulations are plotted in Figures 4, 5, and 6. It can be seen that the sample and asymptotic MISE curves are close to each other, reflecting the relevancy of the first-order asymptotic MISE results. In addition, the minimizers of the two curves are generally close to each other, which bodes well for the adapted plug-in bandwidth selector described in Section 2.2.

5. DATA EXAMPLES

To illustrate the methodology, we use data from the OPEN Study (Kipnis et al. 2003), where we examine body mass index (BMI) distribution in low (<1/3-quantile), moderate (between 1/3-quantile and 2/3-quantile), and high (>2/3-quantile) energy intake groups. The quantiles are the quantiles of usual intake of energy, derived via a measurement error analysis, as described later. Long-term intake of energy is impossible to measure, and so in this case biomarkers with double-labeled water are measured instead. We used log-transformation to make the biomarkers more approximately normally distributed, as was done by Kipnis et al. (2003).

BMI measurements were available for 484 individuals in the sample. Only 24 of these individuals had no biomarker measurements. Thus, we use $\mathbf{q}_i = (1/3, 1/3, 1/3)^{\mathrm{T}}$ for these individuals. The remaining 460 individuals had either one or two measurements of the transformed biomarker, denoted by W_{ij}, $i = 1, \ldots, 460$ and $1 \le j \le r_i$, where r_i is 1 or 2. We assume the standard classical additive error model $W_{ij} = X_i + U_{ij}$, where $X_i \sim \mathrm{Normal}(\mu_x, \sigma_x^2)$ and $U_{ij} \sim \mathrm{Normal}(0, \sigma_u^2)$. Method-of-moments calculations provide estimates of $(\mu_x, \sigma_x^2, \sigma_u^2)$. The sample mean of all the W_{ij} estimates μ_x. To estimate σ_u^2, we take all individuals with $r_i = 2$ replicates, form the differences $W_{i1} - W_{i2}$, and take σ_u^2 as 1/2 the sample variance of the differences. The sample variance of the first biomarkers W_{i1} estimates $\sigma_x^2 + \sigma_u^2$, and thus we estimate σ_x^2 by subtraction. To fully use all of the observations, we now form $W_i = r_i^{-1}\sum_{j=1}^{r_i} W_{ij}$. We thus have $W_i = X_i + U_i$, where $X_i \sim \mathrm{Normal}(\mu_x, \sigma_x^2)$, $U_i \sim \mathrm{Normal}(0, \sigma_{u_i}^2)$, and $\sigma_{u_i}^2 = \sigma_u^2/r_i$. Correspondingly, we have $\sigma_{w_i}^2 = \sigma_x^2 + \sigma_{u_i}^2$. We estimate the probabilities of population membership given the observed biomarker data from normal distribution theory and using the fact that $\mathrm{pr}(X_i \le x | W_i) = \Phi\{(x - \mu_i)/\sigma_i\}$, where $\lambda_i = \sigma_x^2/\sigma_{w_i}^2$, $\mu_i = \mu_x(1 - \lambda_i) + \lambda_i W_i$, and $\sigma_i^2 = \lambda_i \sigma_u^2$.

The estimated probability density functions for the three groups are plotted in Figure 7, based on the three different bandwidth selection methods (ICV, CV, and adapted plug-in). The pointwise 5% and 95% quantile curves are calculated from 1000 bootstrap samples. The two curves in a plot form approximately valid 90% confidence bands for $E\{\widehat{f}_j(x)\}$. Each bootstrap sample has the same q_i values as the original data, whereas the S_i values are generated from the first, second, or third of the three estimated probability density functions depending on whether a variable U_i, randomly generated from uniform [0, 1], is in $[0, q_{i1})$, $[q_{i1}, q_{i1} + q_{i2})$, or $[q_{i1} + q_{i2}, 1]$. Here, when generating the bootstrap samples with the ICV method, the estimated probability density functions are obtained using the bandwidth chosen via ICV. Similarly, when generating the bootstrap samples in the CV or the adapted plug-in method, the estimated probability density functions are obtained using the bandwidth chosen via CV or adapted plug-in. Thus the different bandwidth selection methods not only affect estimation of the density functions in analyzing the original and bootstrap samples, but also cause the bootstrap samples to differ. We can, of course, opt to generate the bootstrap samples from the common estimated density functions, but this seems to be a fairer procedure. Given the similarity between optimal weighted least squares and its

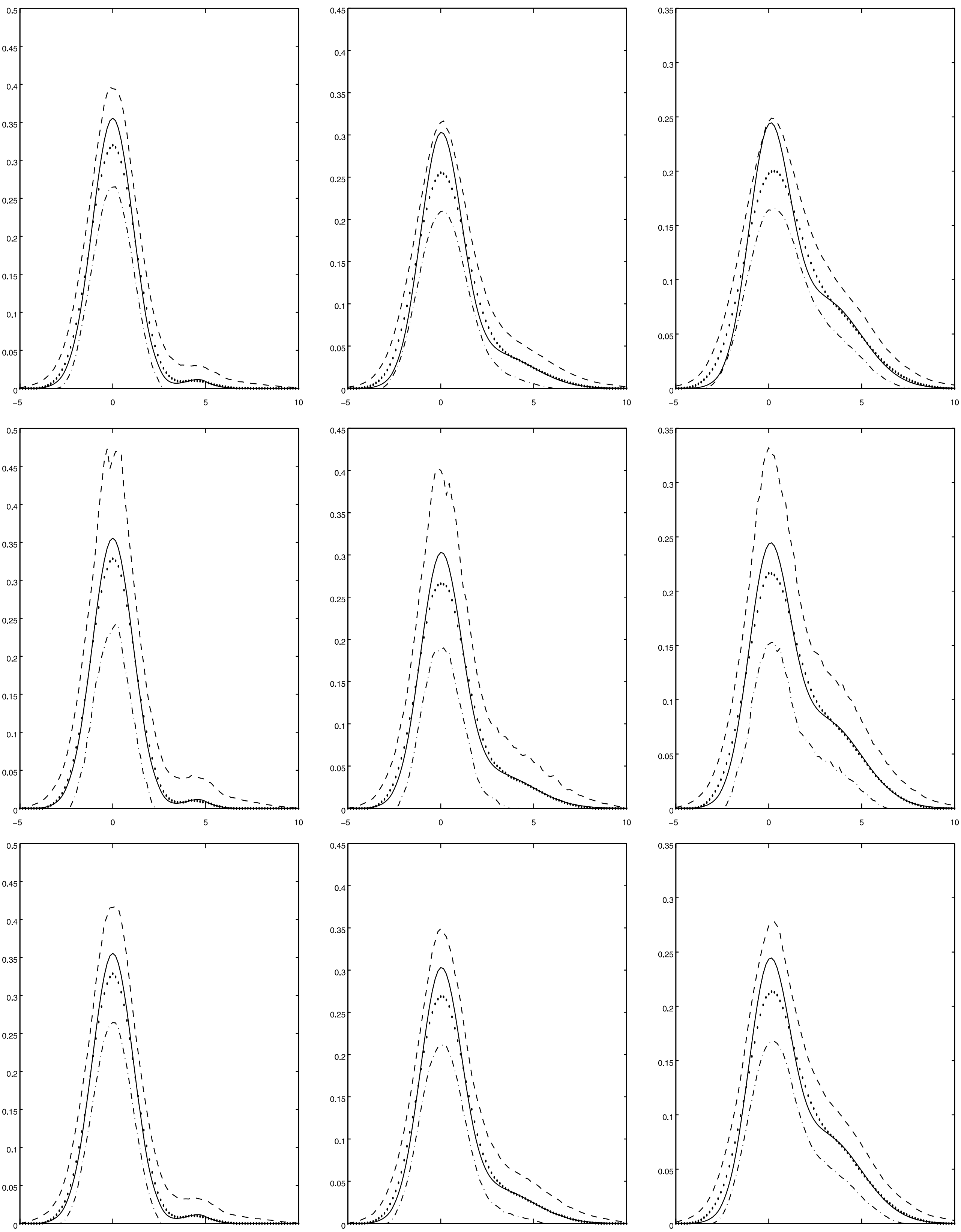

Figure 3. Pointwise quantile curves from simulation 3. In each plot, the solid line is the true density and the other three curves are the median (dotted), 5% (dashed), and 95% (dotted–dashed) quantile curves of all 1000 density estimates in simulation 3. From left to right, the panels correspond to populations 1, 2, and 3, respectively. The rows, from top to bottom, correspond to bandwidth selection by ICV, CV, and the adapted plug-in. All results are based on the constant weight estimators.

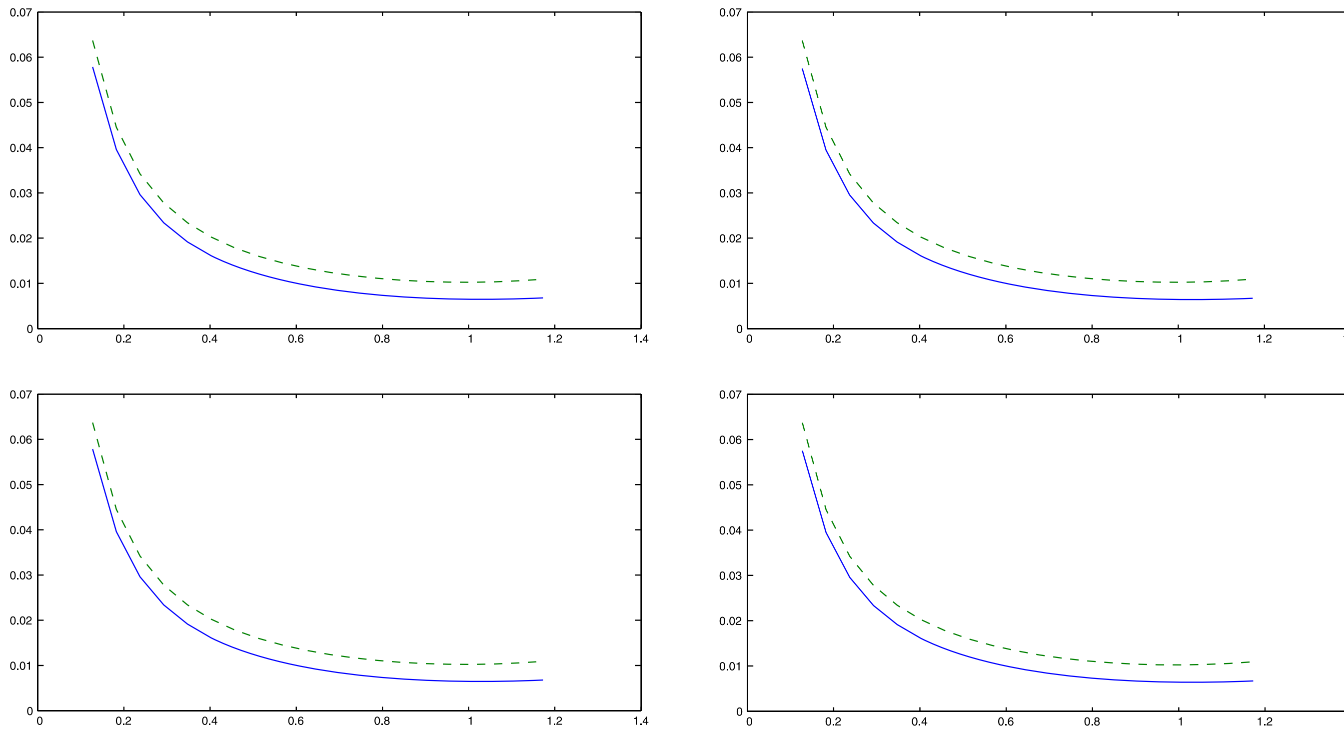

Figure 4. Average ISE (solid) and average of Theorem 1 conditional MISE approximations (dashed) as a function of bandwidth in simulation 1. Results are based on 1000 simulations. The upper and lower panels present OLS and WLS results, respectively. The columns correspond to the different populations. The online version of this figure is in color.

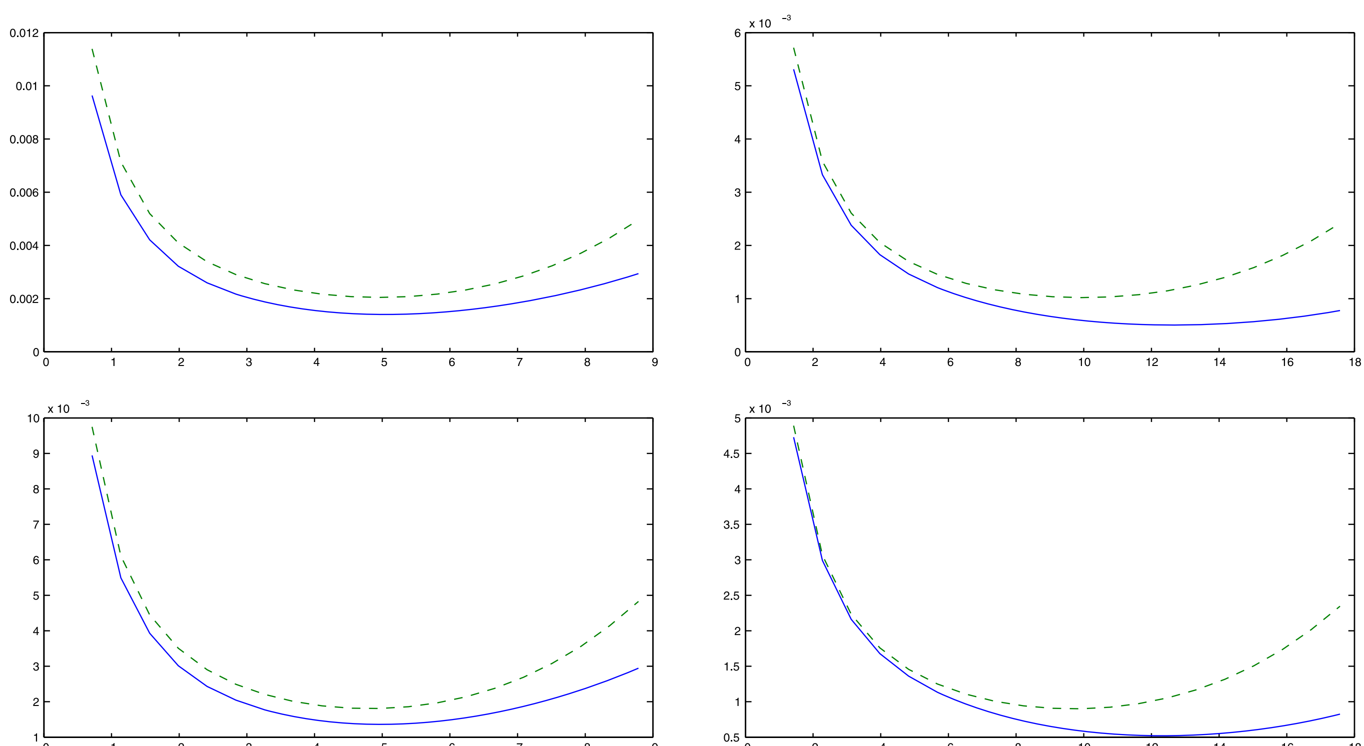

Figure 5. Average ISE (solid) and average of Theorem 1 conditional MISE approximations (dashed) as a function of bandwidth in simulation 2. Results are based on 1000 simulations. The upper and lower panels present OLS and WLS results, respectively. The columns correspond to the different populations. The online version of this figure is in color.

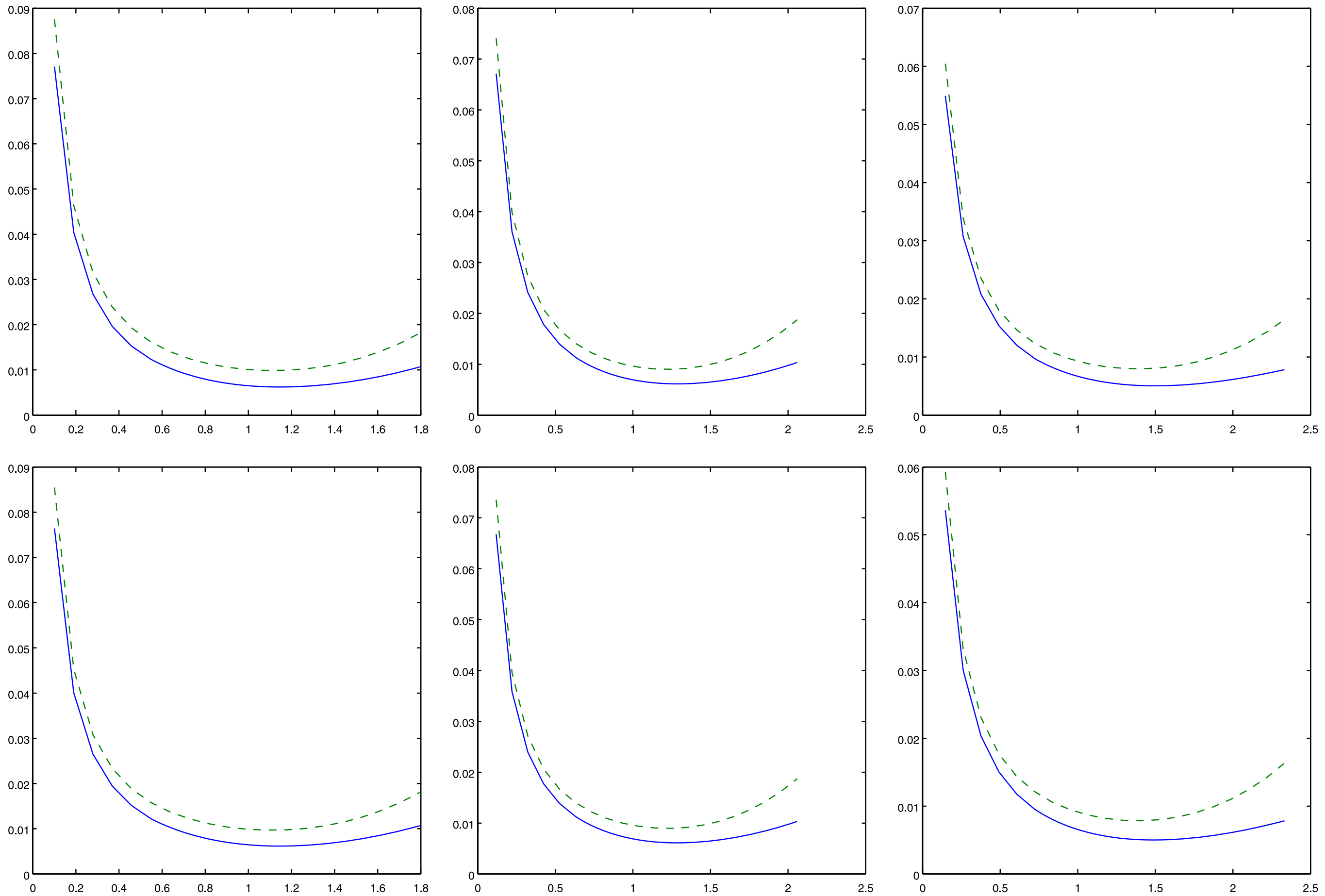

Figure 6. Average ISE (solid) and average of Theorem 1 conditional MISE approximations (dashed) as a function of bandwidth in simulation 3. Results are based on 1000 simulations. The upper and lower panels present OLS and WLS results, respectively. The columns correspond to the different populations. The online version of this figure is in color.

constant weight version exhibited in Section 4, we only implemented the latter. Because of the large variability of the least squares CV, we focus on the results from the ICV and adapted plug-in methods. Although the estimated density functions are slightly different among the methods, an obvious conclusion common to both methods is that there is a continuing shift of mass from the small BMI region to the large BMI region from the first population to the third population, verifying intuition that higher energy intake can lead to increased BMI.

The variability increases from population 1 to population 2 to population 3, and this is especially evident from population 2 to population 3. Finally, each intake group seems to have a small subgroup in which the BMI values form a small mass to the right of the majority of that group. The size of this subgroup increases as consumption increases. This indicates that a small number of individuals tend to have very high BMI (BMI > 35, classified as class II obese) even if their energy intake is relatively low.

6. DISCUSSION

For simplicity of implementation, here we have provided a one-step density estimation procedure. The trade-off of doing so is that the estimated density functions in finite samples are not guaranteed to be nonnegative and to integrate to 1. To overcome this, we can of course perform some simple postestimation adjustment, for example, set the negative values to 0 and scale each density so that it integrates to 1. In practice, as long as the sample size is not too small, this postestimation procedure causes only a slight change, while improving such features as MISE.

If nonnegativity and integration to 1 are of essential importance, then we also can construct more complex estimators. For example, let $\widetilde{\mathbf{f}}(s)$ be an initial nonnegative estimate of $\mathbf{f}(s)$, and let

$$a_{ij}(u) = \frac{\widetilde{f}_j(u) q_{ij}}{\mathbf{q}_i^{\mathrm{T}} \widetilde{\mathbf{f}}(u)}.$$

Then we can update the estimate of $f_j(s), j = 1, \ldots, P$, via

$$\widetilde{f}_j(s) = \sum_{i=1}^{n} \frac{a_{ij}(S_i)}{\sum_{k=1}^{n} a_{kj}(S_k)} K\left(\frac{s - S_i}{h}\right).$$

This procedure can be iterated until convergence to obtain $\widehat{\mathbf{f}}(s)$. It can be readily seen that $\widetilde{f}_j(s)$ is always nonnegative and integrates to 1. However, because in each update we need to select a corresponding bandwidth, the required computation is extensive.

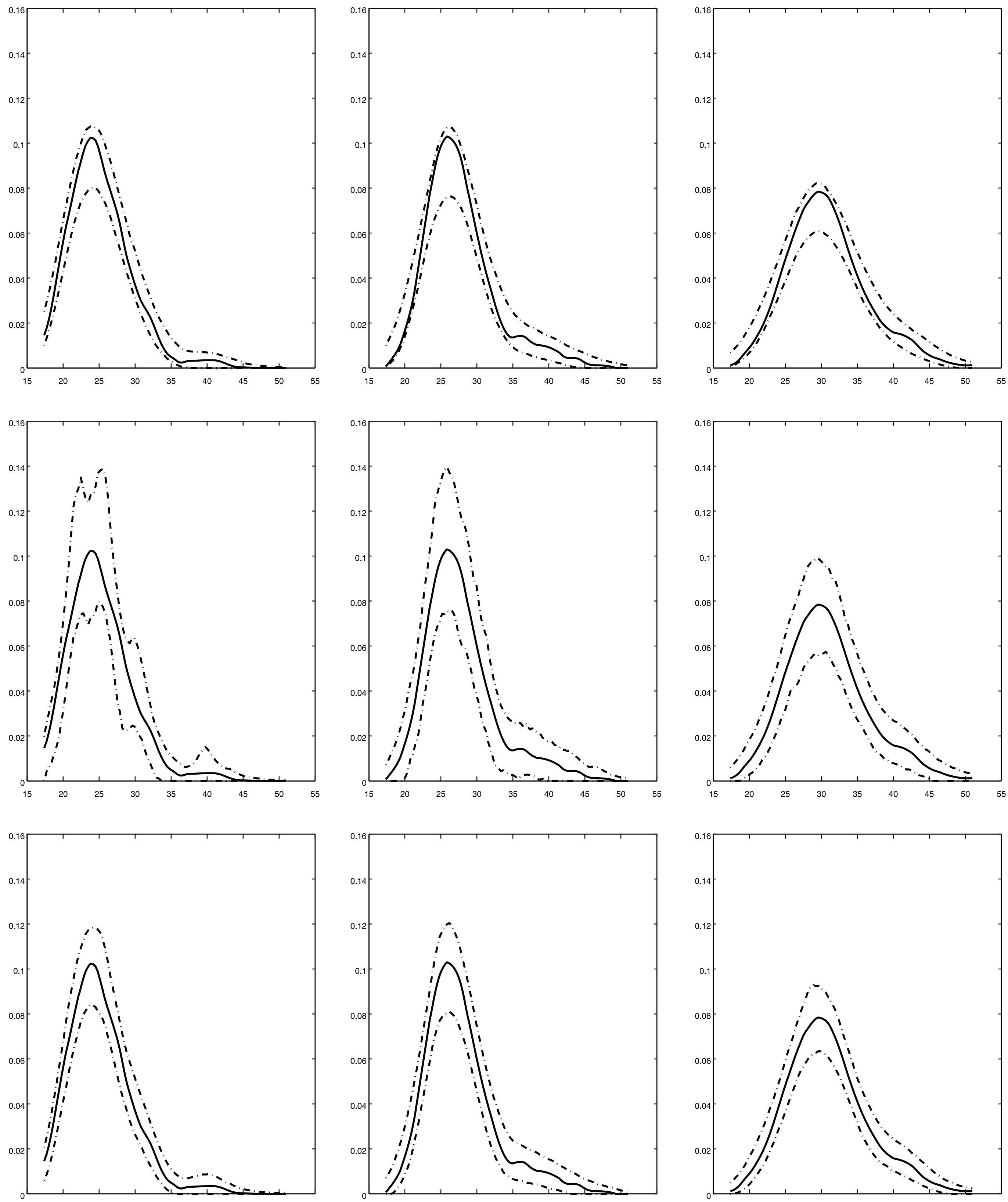

Figure 7. Density estimates and 90% bootstrap confidence bands for the BMI data. The left, middle, and right panels correspond to first, second, and third tertiles, respectively, of the energy intake population. The rows correspond to bandwidth selection via ICV (top), CV (middle), and the adapted plug-in method (bottom), all based on the constant weight estimators.

APPENDIX

A.0 List of Regularity Conditions

C1. K is a symmetric second-order kernel that is bounded in absolute value and has compact support.

C2. For $j = 1, \dots, P$, $h_j \to 0$ and $nh_j^3 \to \infty$ as $n \to \infty$.

C3. Each f_j and each proposed density function f_j^* are twice differentiable with bounded second derivative. Each f_j^* is bounded and strictly positive throughout a neighborhood of s, and the variance of estimator (8) exists and is finite for each $j = 1, \dots, P$.

A.1 Proof of Theorem 1

The bias of $\widehat{f}_j(s)$ is

$$
\begin{aligned}
E\{\widehat{f}_j(s)\} - f_j(s) &= \mathbf{e}_j^{\mathrm{T}} \left(n^{-1} \sum_{i=1}^n w_i \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}} \right)^{-1} \\
&\quad \times n^{-1} \sum_{i=1}^n w_i \mathbf{q}_i E K_{h_j}(S_i - s) - f_j(s) \\
&= \mathbf{e}_j^{\mathrm{T}} \left(n^{-1} \sum_{i=1}^n w_i \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}} \right)^{-1} \\
&\quad \times n^{-1} \sum_{i=1}^n w_i \mathbf{q}_i \int K_{h_j}(t - s) \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(t)\, dt - f_j(s) \\
&= \mathbf{e}_j^{\mathrm{T}} \left(n^{-1} \sum_{i=1}^n w_i \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}} \right)^{-1} \\
&\quad \times n^{-1} \sum_{i=1}^n w_i \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}} \int K_{h_j}(t - s) \mathbf{f}(t)\, dt - f_j(s) \\
&= \int K_{h_j}(t - s) f_j(t)\, dt - f_j(s) \\
&= h^2 f_j''(s) \sigma_K^2 / 2 + O(h^4).
\end{aligned}
$$

The variance of $\widehat{f}_j(s)$ is

$$
\begin{aligned}
\mathrm{var}\{\widehat{f}_j(s)\} &= \frac{\mathbf{e}_j^{\mathrm{T}}}{n^2} \mathcal{M}_q \sum_{i=1}^n w_i^2 \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}} \\
&\quad \times \left[E K_{h_j}^2(S_i - s) - \{E K_{h_j}(S_i - s)\}^2 \right] \mathcal{M}_q \mathbf{e}_j \\
&= \mathbf{e}_j^{\mathrm{T}} \mathcal{M}_q \frac{1}{n^2} \sum_{i=1}^n w_i^2 \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}} \left\{ \frac{\mathbf{q}_i^{\mathrm{T}} \mathbf{f}(s)}{h_j} C_K + O(1) \right\} \mathcal{M}_q \mathbf{e}_j \\
&= \frac{1}{nh_j} \mathbf{e}_j^{\mathrm{T}} \mathcal{M}_q \left\{ n^{-1} \sum_{i=1}^n w_i^2 \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}} \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(s) \right\} \mathcal{M}_q \mathbf{e}_j C_K + O(n^{-1}).
\end{aligned}
$$

A.2 Proof of Theorem 2

Let the estimate from (1) with the weights $w_i = \{\mathbf{q}_i^{\mathrm{T}} \mathbf{f}^*(s)\}^{-1}$ be $\widetilde{f}_j$, and let the estimate from (8) be $\widehat{f}_j$. We can write $\widetilde{f}_j(s) - \widehat{f}_j(s) = n^{-1} \sum_{i=1}^n d_i(s)$, where for $i = 1, \dots, n$,

$$
\begin{aligned}
d_i(s) &= \mathbf{e}_j^{\mathrm{T}} \left(n^{-1} \sum_{i=1}^n \frac{\mathbf{q}_i \mathbf{q}_i^{\mathrm{T}}}{\mathbf{q}_i^{\mathrm{T}} \mathbf{f}^*(s)} \right)^{-1} \frac{\mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}} \mathbf{f}^*(s)} K_{h_j}(S_i - s) \\
&\quad - \mathbf{e}_j^{\mathrm{T}} \frac{\mathbf{A}^*(S_i)^{-1} \mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}} \mathbf{f}^*(S_i)} K_{h_j}(S_i - s) \\
&= \mathbf{e}_j^{\mathrm{T}} \left\{ \frac{\mathbf{A}^*(s)^{-1} \mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}} \mathbf{f}^*(s)} - \frac{\mathbf{A}^*(S_i)^{-1} \mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}} \mathbf{f}^*(S_i)} \right\} K_{h_j}(S_i - s) \\
&= u(\mathbf{q}_i, S_i, s) K_{h_j}(S_i - s)
\end{aligned}
$$

and

$$
u(\mathbf{q}_i, S_i, s) = \mathbf{e}_j^{\mathrm{T}} \left\{ \frac{\mathbf{A}^*(s)^{-1} \mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}} \mathbf{f}^*(s)} - \frac{\mathbf{A}^*(S_i)^{-1} \mathbf{q}_i}{\mathbf{q}_i^{\mathrm{T}} \mathbf{f}^*(S_i)} \right\}.
$$

Note that $n^{-1} \sum_{i=1}^n u(\mathbf{q}_i, t, s) \mathbf{q}_i^{\mathrm{T}} = 0$ for any s, t and $u(\mathbf{q}_i, s, s) = 0$ for all s. The first of these properties yields

$$
\begin{aligned}
E\left\{ n^{-1} \sum_{i=1}^n d_i(s) \right\} &= n^{-1} \sum_{i=1}^n \int u(\mathbf{q}_i, t, s) K_{h_j}(t - s) \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(t)\, dt \\
&= \int n^{-1} \sum_{i=1}^n u(\mathbf{q}_i, t, s) \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(t) K_{h_j}(t - s)\, dt = 0.
\end{aligned}
$$

Furthermore, we obtain

$$
\begin{aligned}
&E\left[\left\{ n^{-1} \sum_{i=1}^n d_i(s) \right\}^2 \right] \\
&\quad = E\left[\left\{ n^{-1} \sum_{i=1}^n u(\mathbf{q}_i, S_i, s) K_{h_j}(S_i - s) \right\}^2 \right] \\
&\quad = \frac{1}{n^2} \sum_{i,k=1, i \neq k}^n \int u(\mathbf{q}_i, t, s) K_{h_j}(t - s) \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(t)\, dt \\
&\qquad \times \int u(\mathbf{q}_k, t, s) K_{h_j}(t - s) \mathbf{q}_k^{\mathrm{T}} \mathbf{f}(t)\, dt \\
&\qquad + \frac{1}{n^2} \sum_{i=1}^n \int u^2(\mathbf{q}_i, t, s) K_{h_j}^2(t - s) \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(t)\, dt \\
&\quad = \frac{1}{n^2} \sum_{i=1}^n \int u(\mathbf{q}_i, t, s) K_{h_j}(t - s) \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(t)\, dt \\
&\qquad \times \sum_{k=1}^n \int u(\mathbf{q}_k, t, s) K_{h_j}(t - s) \mathbf{q}_k^{\mathrm{T}} \mathbf{f}(t)\, dt \\
&\qquad + \frac{1}{n^2} \sum_{i=1}^n \int u^2(\mathbf{q}_i, t, s) K_{h_j}^2(t - s) \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(t)\, dt \\
&\qquad - \frac{1}{n^2} \sum_{i=1}^n \left\{ \int u(\mathbf{q}_i, t, s) K_{h_j}(t - s) \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(t)\, dt \right\}^2 \\
&\quad = 0 + \frac{1}{n^2 h_j} \sum_{i=1}^n \int u^2(\mathbf{q}_i, s + h_j u, s) K^2(u) \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(s + h_j u)\, du \\
&\qquad - \frac{1}{n^2} \sum_{i=1}^n \left\{ \int u(\mathbf{q}_i, s + h_j u, s) K(u) \mathbf{q}_i^{\mathrm{T}} \mathbf{f}(s + h_j u)\, du \right\}^2 \\
&\quad = O(n^{-1} h_j).
\end{aligned}
$$

The last result follows from using condition C2 and $u(\mathbf{q}_i, s, s) = 0$.

[Received December 2010. Revised April 2011.]

REFERENCES

Bowman, A. W. (1984), "An Alternative Method of Cross-Validation for the Smoothing of Density Estimates," *Biometrika*, 71, 353–360. [1181]

Dodd, K., Guenther, P. M., Freedman, L. S., Subar, A. F., Kipnis, V., Midthune, D., Tooze, J. A., and Krebs-Smith, S. M. (2006), "Statistical Methods for Estimating Usual Intake of Nutrients and Foods: A Review of the Theory," *Journal of the American Dietetic Association*, 106, 1640–1650. [1180]

Johnson, B. A., Herring, A. H., Ibrahim, J. G., and Siega-Riz, A. M. (2007), "Structured Measurement Error in Nutritional Epidemiology: Applications in the Pregnancy, Infection, and Nutrition (PIN) Study," *Journal of the American Statistical Association*, 102, 856–866. [1180]

Kipnis, V., Subar, A. F., Midthune, D., Freedman, L. S., Ballard-Barbash, R., Troiano, R., Bingham, S., Schoeller, D. A., Schatzkin, A., and Carroll, R. J. (2003), "The Structure of Dietary Measurement Error: Results of the OPEN Biomarker Study," *American Journal of Epidemiology*, 158, 14–21. [1186]

Ma, Y., and Wang, Y. (2010), "Efficient Semiparametric Estimation for Mixture Data," manuscript, Texas A&M University and Columbia University. [1180-1183,1186]

Mallapragada, P. K., Jin, R., and Jain, A. (2010), "Non-Parametric Mixture Models for Clustering," in *Proceedings of the 2010 Joint IAPR International Conference on Structural, Syntactic, and Statistical Pattern Recognition*, Berlin/Heidelberg: Springer, pp. 334–343. [1180]

Nusser, S. M., Fuller, W. A., and Guenther, P. M. (1997), "Estimating Usual Dietary Intake Distributions: Adjusting for Measurement Error and Nonnormality in 24-Hour Food Intake Data," in *Survey Measurement and Process Quality*, eds. L. Lyberg, P. Biemer, M. Collins, E. Deleeuw, C. Dippo, N. Schwartz, and D. Trewin, New York: Wiley, pp. 670–689. [1180]

Savchuk, O. Y., Hart, J. D., and Sheather, S. J. (2010), "Indirect Cross-Validation for Density Estimation," *Journal of the American Statistical Association*, 105, 415–423. [1181]

Sheather, S. J., and Jones, M. C. (1991), "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation," *Journal of the Royal Statistical Society, Ser. B*, 53, 683–690. [1182]

Sinha, S., Mallick, B. K., Kipnis, V., and Carroll, R. J. (2010), "Semiparametric Bayesian Analysis of Nutritional Epidemiology Data in the Presence of Measurement Error," *Biometrics*, 66, 444–454. [1180]